Single-cell analysis of neurodevelopment in Huntington's disease

Jordi Abante[1-3], Cristina Vila[1], Olga Varea[1], Petia I. Radeva[2], Josep Mª Canals[1]

[1] Laboratory of Stem Cells and Regenerative Medicine, Department of Biomedical Sciences; Creatio - Production and Validation Center of Advanced Therapies, Faculty of Medicine and Health Sciences; Institute of Neurosciences, University of Barcelona; August Pi i Sunyer Biomedical Research Institute (IDIBAPS), Barcelona, 08036, Spain.

[2] Department of Mathematics and Computer Science, University of Barcelona, Barcelona, 08036, Spain

[3] Department of Biomedical Data Science, Stanford University, Stanford, 94305, USA

The study of the embryonic development of specific regions of the brain can provide valuable insights into the molecular mechanisms that underlie their formation and function, with potential implications for neurodegenerative disease research. However, the development of some regions, such as the striatum, is yet to be fully characterized. Single-cell and single-nuclei RNA sequencing technologies offer a great opportunity to study developmental transcriptomic events of this region, which plays a critical role in Huntington's disease (HD). Here we develop a computational pipeline to integrate single-cell and single-nuclei RNA sequencing data at different developmental stages to gain insights into alterations during development in HD.

Obtaining enough cells for sequencing from such regions in embryonic stages can be challenging and may require pooling tissue from multiple animals of different sexes adding variability to the sequencing data. We propose a probabilistic approach to determine the sex of the individual cells in mixed populations, enabling the inclusion of this variable in downstream modeling. To integrate the different modalities, we explicitly correct for technical biases in the data and use canonical correlation analysis to find a common representation that maximizes the covariance across datasets. Upon integration of the data, we perform various analyses, including composition analysis to determine the enrichment or depletion of certain cell types as well as differential expression analysis to determine genes that are dysregulated in HD. For the former, we develop a composition analysis framework based on generalized linear models that allows us to properly control for possible confounders.

This work demonstrates the utility of computational analysis of transcriptomics data to gain insights into alterations during embryonic development in HD. The proposed pipeline can be applied to other brain regions and developmental stages to provide further insights into the molecular mechanisms underlying neurodevelopmental and other neurodegenerative diseases.

Funding: This study was supported by grants from the Ministerio de Ciencia e Innovación (PID2021-126961OB-I00); Instituto de Salud Carlos III, Ministerio de Ciencia e Innovación and European Regional Development Fund (ERDF) (Red de Terapias Avanzadas, RD21/0017/0020); and Generalitat de Catalunya (2021 SGR 01094), Spain; and "la Caixa" Foundation under the grant agreement LCF/PR/HR21-00622".